This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.3009903, IEEE Journal of Biomedical and Health Informatics

Estimation of Instantaneous Oxygen Uptake during Exercise and Daily Activities using a Wearable Cardio-Electromechanical and Environmental Sensor

Md Mobashir Hasan Shandhi, *Student Member, IEEE,* William H. Bartlett, J. Alex Heller, Mozziyar Etemadi, Aaron Young, Thomas Plötz, and Omer T. Inan, *Senior Member, IEEE*

Abstract*—*Objective: **To estimate instantaneous oxygen uptake (VO_2) with a small, low-cost wearable sensor during exercise and daily activities in order to enable monitoring of energy expenditure (EE) in uncontrolled settings. We aim to do so using a combination of seismocardiogram (SCG), electrocardiogram (ECG) and atmospheric pressure (AP) signals obtained from a minimally obtrusive wearable device.** ***Methods:*** **In this study, subjects performed a treadmill protocol in a controlled environment and an outside walking protocol in an uncontrolled environment. During testing, the COSMED K5 metabolic system collected gold standard breath-by-breath (BxB) data and a custom-built wearable patch placed on the mid-sternum collected SCG, ECG and AP signals. We extracted features from these signals to estimate the BxB VO_2 data obtained from the COSMED system.** ***Results:*** **In estimating instantaneous VO_2, we achieved our best results on the treadmill protocol using a combination of SCG (frequency) and AP features (RMSE of 3.68$\pm$0.98 ml/kg/min and R^2 of 0.77). For the outside protocol, we achieved our best results using a combination of SCG (frequency), ECG and AP features (RMSE of 4.3$\pm$1.47 ml/kg/min and R^2 of 0.64). In estimating VO_2 consumed over one minute intervals during the protocols, our median percentage error was 15.8% for the treadmill protocol and 20.5% for the outside protocol.** ***Conclusion:*** **SCG, ECG and AP signals from a small wearable patch can enable accurate estimation of instantaneous VO_2 in both controlled and uncontrolled settings. SCG signals capturing variation in cardio-mechanical processes, AP signals, and state of the art machine learning models contribute significantly to the accurate estimation of instantaneous VO_2.** ***Significance:*** **Accurate estimation of VO_2 with a low cost, minimally obtrusive wearable patch can enable the monitoring of VO_2 and EE in everyday settings and make the many applications of these measurements more accessible to the general public.**

***Index Terms*—Seismocardiography, COSMED K5, Oxygen Uptake, Metabolic, Machine Learning**

I. Introduction

Oxygen uptake (VO_2) is an important physiological parameter for numerous reasons. VO_2 max, an individual's maximum rate of oxygen uptake during incremental exercise, is commonly used as a measurement of cardiorespiratory fitness (CRF) for both clinical and non-clinical purposes. Athletes frequently measure their VO_2 max to assess their endurance. Medical professionals measure CRF with cardiopulmonary exercise testing (CPET) to quantify disease progression in patients with conditions such as heart failure (HF) and chronic obstructive pulmonary disease [1]–[4]. Submaximal VO_2 measurement also has many applications. It can be used along with a subject's weight to calculate energy expenditure, a critical ingredient in the calculation of energy balance, which quantifies the relationship between an individual's caloric intake and expenditure. Energy balance is important for the management of many illnesses such as diabetes, obesity, cardiovascular disease, and cancer, as well as for maintaining a generally healthy lifestyle [5], [6].

Measuring VO_2 is largely limited to clinical environments due to extensive equipment requirements and the need for professionally trained technicians. The ability to test more ubiquitously could enable many desirable applications such as more consistent fitness tracking, closer monitoring of disease progression, and more precise weight-loss management. To this aim, companies such as COSMED, VO2 Master, and VacuMed have developed portable calorimetry systems capable of providing accurate VO_2 measurements in outdoor environments [7], [8]. However, these systems are both prohibitively expensive for use by the general public and still involve highly obtrusive components (e.g. masks etc), making them unsuitable for ubiquitous monitoring. For this reason, there have been many attempts to achieve similar readings from minimally obtrusive wearable devices [9]. Commercially available examples include FitBit, Apple Watch, and Samsung Gear S2. However, studies have repeatedly shown that while these devices produce accurate heart rate (HR) measurements in laboratory settings, their EE measurements do not meet acceptable standards [10]–[14]. These studies warn against using any of these commercial systems for weight or disease management programs.

Many research efforts have thus sought to obtain superior portable VO_2 estimation from minimally obtrusive wearable devices. As far back as 1981 [15], studies have demonstrated the ability to obtain reasonable estimations of EE with various methods involving accelerometry [15]–[17] and, more recently, gyroscopes [17]. However, accelerometry-based approaches often err when motion does not involve the body part where the

Research reported in this publication was supported, in part, by the National Heart, Lung and Blood Institute under R01HL130619. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

M. M. H. Shandhi and O. T. Inan are with School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. Corresponding author email: mobashir.shandhi@gatech.edu

W. H. Bartlett and T. Plötz are with School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, USA.

J. A. Heller and M. Etemadi are with Feinberg School of Medicine, Northwestern University, Chicago, IL, USA.

A. Young is with Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

sensor is placed, causing accuracy to vary significantly according to activity and/or placement location [9], [18]. As a result, other recent efforts have demonstrated improved results combining accelerometry with HR and electrocardiogram (ECG) sensors to incorporate cardiac information [19]. Another study showed the potential of utilizing an additional physiological measurement–ventilation–to produce accurate estimations of VO_2 [20].

While most studies were conducted only in laboratory settings with protocols involving treadmill-based exercise and/or standardized activities, several studies of note showed results in free-living conditions. Participants in a study by Lester et al. conducted a field protocol involving stairs, inclines, elevators, object lifting, and sweeping, both indoors and outdoors [21]. While this study, which incorporates grade using barometric pressure and GPS, achieved considerable accuracy, the use of regression formulas developed by the American College of Sports Medicine (ACSM) necessitates the restrictive and error-prone sub-task of five-class activity recognition. Additionally, in two separate studies, Bouarfa et al. [22] and Brage et al. [23] achieved low error in estimating EE over 14 days of free-living [24]. Bouarfa et al. [22] used a single ear-worn sensor and Brage et al. [23] used the combined HR and accelerometry Actiheart system, both comparing their results against ground truth from doubly-labeled water. However, neither study could assess instantaneous results as their target value for each subject was a single energy expenditure reading totaled over the full trial period.

Our study introduces the combination of state-of-the-art machine learning algorithms with seismocardiogram (SCG)–the measure of thoracic vibrations produced by cardio-ventricular contraction and blood ejection into the vascular tree [25]–for estimating instantaneous VO_2 in indoor and outdoor settings. Recent studies [26]–[28] involving patients with HF have shown that VO_2 max and the clinical status of a patient with HF can be estimated from the electrocardiogram (ECG) and SCG signals using a small wearable patch. All of these studies were conducted with HF patients in controlled clinical settings by trained professionals. In this study, rather than obtaining the singular VO_2 max parameter in clinical settings for disease monitoring purposes, we seek to estimate instantaneous sub-maximal VO_2 in healthy individuals in both controlled and uncontrolled environments with advanced machine learning algorithms.

We fitted subjects with two simultaneously recording data collection systems in this study. We placed a small wearable patch that collects ECG, SCG, and atmospheric pressure (AP) measurements on a subject's mid-sternum. This patch is an improvement upon our previous version [29]. A triaxial accelerometer contained within the patch measures SCG. We measure gold standard VO_2 using the COSMED K5 wearable metabolic system. While previous studies measuring EE and VO_2 with single wearable sensors have also used accelerometry-based methods, the accelerometer is commonly placed in distal locations such as the waist [19], [21], [30] or ear [22]. In our case, we place an accelerometer on the chest to capture SCG signals that specifically relate to cardio-mechanical activity in ways that remotely placed accelerometers are not able to. As a result, we avoid the problem of having to determine placement location and potentially reduce variability across activities. We compare the efficacy of different environmental, ECG, and SCG-based feature sets generated from the wearable patch in combination with models of varying levels of complexity to explore how these different types of signals relate to oxygen uptake during physical activity, ultimately arriving at a global regression model.

II. Methods

A. Experimental Protocol

We conducted this study under a protocol (H18452) approved by the Georgia Institute of Technology Institutional Review Board. A total of 17 healthy subjects (9 females and 8 males) participated in the study (Age: 26.8 $\pm$ 4.1 years, Weight: 67.5 $\pm$ 14.1 kg and Height: 170.5 $\pm$ 9.9 cm). All subjects provided written informed consent before experimentation and reported no cardiopulmonary issues.

Fig. 1(a) shows the placement of both sensors: our custom-built wearable patch and the COSMED K5 (COSMED, Rome, Italy) metabolic system. Fig. 1(b) shows the custom-built wearable sensor hardware, which measures ECG, triaxial SCG, and environmental features (atmospheric pressure, temperature and humidity). For each subject, we placed the wearable sensor evenly between the suprasternal notch and xiphoid process on the mid-sternal line, using three ECG electrodes (model 2670, 3M, Saint Paul, MN, USA). For the COSMED K5 system, we fitted subjects with a gas exchange mask on their face and the K5 system on their back. We situated a heart rate belt from the K5 system just below the chest line. After fitting subjects with all the sensors and systems, we asked them for confirmation of their comfort before testing. At the start of each trial, we synchronized both the wearable sensor and K5 system to a smart mobile phone in order to record timestamps throughout the protocol.

Fig. 1(d) shows the outline of the study protocol, which we divided into two parts: a treadmill walking portion in a laboratory setting and an outdoor walking portion in an uncontrolled setting. For the treadmill part of the protocol, subjects first stood still for two minutes to record baseline data. Then, subjects walked on a treadmill at five different speed settings (0.75, 1, 1.25, 1.5 and 1.75 meters/second, roughly 1.7, 2.2, 2.8, 3.4 and 4 miles/hour) for six minutes at each speed, totaling 30 minutes of walking. After completing the treadmill walk, subjects stood still for 5 minutes to record a recovery period. Including this recovery period, subjects performed the treadmill protocol continuously for a total of 37 minutes.

After the treadmill part, subjects rested for 15 minutes before starting the outside walking protocol. They began this section standing still for 2 minutes at the "Start/Stop" location of the route marked in Fig. 1(c). Then, subjects walked the route shown in Fig. 1(c), arriving back at the same "Start/Stop" point. The route contains a mixture of level ground, uphill and downhill walkways (with significant slopes), two uphill stairs climbing and four traffic signals along the way. The terrain is marked in Fig. 1(c). Subjects completed this walk at their own chosen speed and followed normal pedestrian traffic

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.3009903, IEEE Journal of Biomedical and Health Informatics

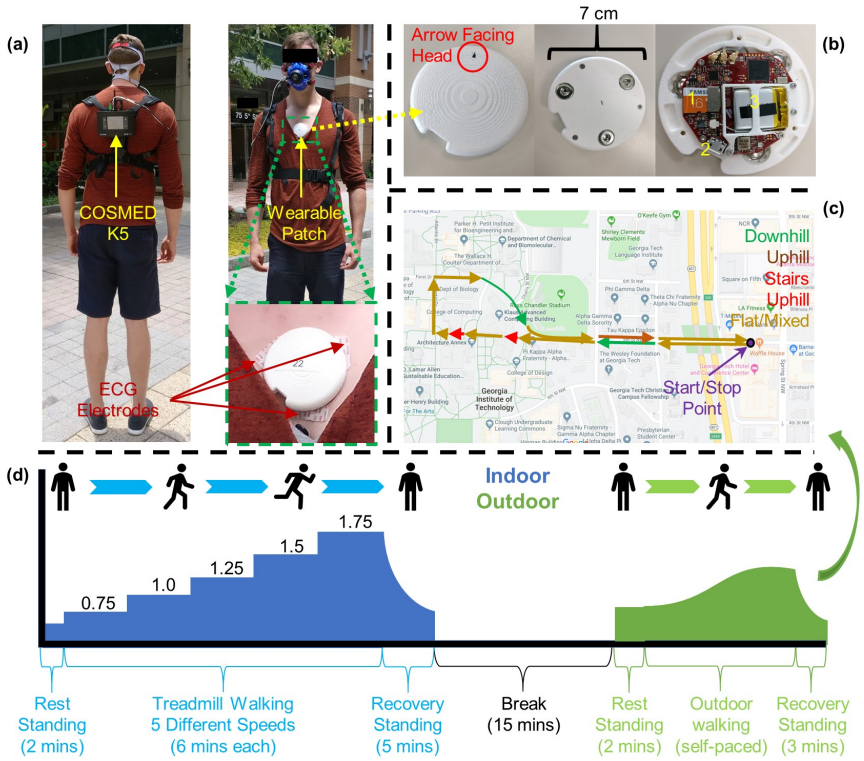

Fig. 1. (a) A subject configured with both the wearable patch and the COSMED K5 system, with inset showing a zoomed-in image of the ECG electrodes. (b) The wearable patch top, bottom, and inside. The microSD (1), microUSB (2), and battery (3) are shown. (c) A map of the outdoor walking route across Georgia Institute of Technology with marked terrain. (d) An outline of the study protocol.

laws. After completing the route, subjects stood still for 3 minutes to record a recovery period. This part of the protocol took approximately 20-30 minutes depending on each subject's speed as well as traffic conditions during their testing. Of 17 subjects, two females were not able to perform the outside walk due to precipitation. For that reason, we obtained data from 17 treadmill protocols and 15 outdoor protocols.

B. Sensing Hardware

We recorded gold standard breath-by-breath (BxB) metabolic data with the COSMED K5 system (COSMED, Rome, Italy). Subjects wore the COSMED heart rate probe (i.e. belt) which provided a HR reading corresponding to the ground-truth metabolic data.

We collected ECG and triaxial SCG (axes: head-to-foot (HtoF), dorso-ventral (DV), and lateral (Lat)), with a novel wearable patch as shown in Fig. 1(b). This patch is an improvement upon our previous version as described in [29]. It contains an ATSAM4LS microcontroller (Atmel Corporation, San Jose, CA), whereas the previous version used an ATmega1284P microcontroller (Microchip Technology, Chandler, AZ). The ECG sensor uses an analog-front-end-integrated circuit with an on-board analog-to-digital converter ADS1291 (Texas Instruments, Dallas, TX). The accelerometer in the present patch that acquires triaxial SCG signals is the ADXL355 (Analog Devices, Norwood, MA), which has a low noise floor of 25 $\mu g/\sqrt{Hz}$ compared to the triaxial accelerometer BMA280 (Bosch Sensortech GmbH, Reutlingen, Germany) used in the previous version [29] with a noise floor of 120 $\mu g/\sqrt{Hz}$. The patch also contains a BME280 (Bosch Sensortech GmbH, Reutlingen, Germany) environmental sensor which records atmospheric pressure (AP), ambient temperature and relative humidity, whereas the previous version [29] had only pressure sensing capability using the MS5611-01BA03 (Measurement Specialties, Fremont, CA). The patch used in this work has a diameter of 7 cm and weight of 38.2 gm. When fully charged, it can record continuously for approximately 45 hours, which is more than sufficient for constant remote monitoring. Initially it samples the ECG signal at 1kHz, the accelerometer signals at 500 Hz and the environmental signals at 20 Hz, and saves the data into a SD card in the patch. A custom-built graphical user interface accesses all the data into a computer and resamples the accelerometer and environmental signals at 1 kHz to have the same sampling frequency of 1 kHz for all the signals.

C. Signal Processing and Feature Extraction

1) Pre-processing and Linear Filtering

To assess VO_2 estimation ability in controlled and uncontrolled settings independently, data from the treadmill and outside protocols were processed and analyzed separately. Later, the models trained on the data from one protocol were validated on the data from the other protocol and vice versa, to evaluate the generalizability of the models.

Fig. 2(a) illustrates the signal processing and feature extraction procedures used for the wearable signals (ECG, SCG and AP) and the BxB COSMED K5 data. We synchronized all signals from the wearable patch with the BxB data from the COSMED K5 system and took relevant timestamps with a mobile phone to be used in subsequent analysis. We filtered

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.3009903, IEEE Journal of Biomedical and Health Informatics

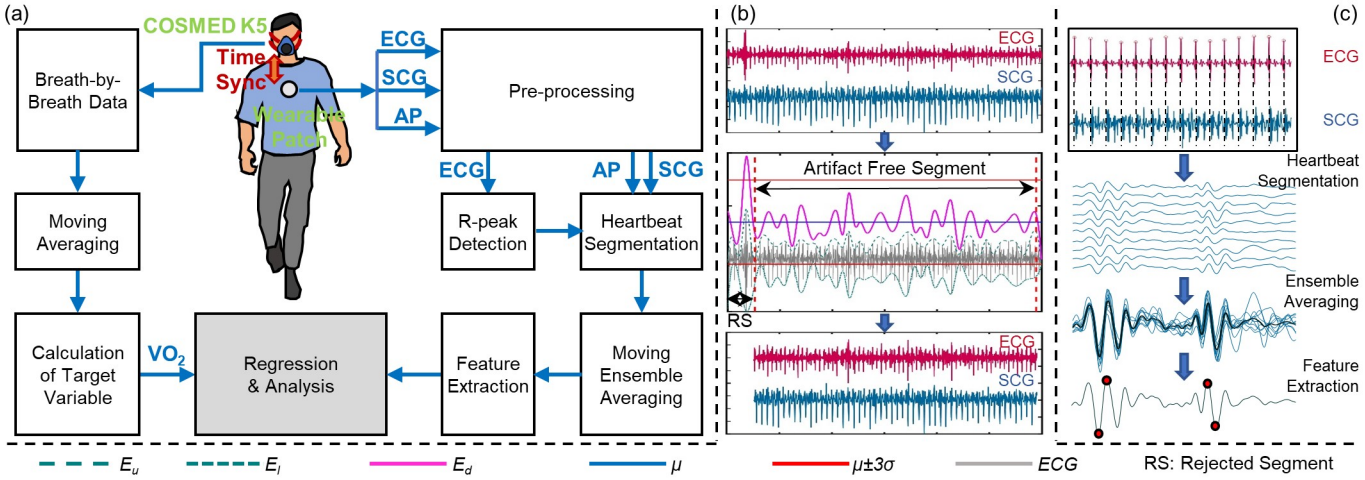

Fig. 2. (a) Functional block diagram of signal processing and feature extraction procedure: Wearable ECG, SCG and AP signals were synchronized with breath-by-breath (BxB) data from COSMED K5 system. Wearable ECG and SCG signals were filtered and sections corrupted with motion artifacts were removed from the signals. The ECG signal was normalized and the R-peaks of the ECG signal were detected, and the SCG and AP signals were segmented into heartbeats using corresponding R-peaks. SCG and AP heartbeats were averaged to get ensemble averaged heartbeats and features were extracted from the averaged beats. The wearable features were fed into regressor as estimators. VO_2 from the BxB data corresponding to each averaged heartbeats were computed and fed into regressor as target variables. (b) Visual representation of ECG artifact removal (with legends showing different lines for the figure in the middle). (c) Visual representation of heartbeat segmentation, ensemble averaging and feature extraction of SCG_{DV} signal.

The raw ECG and SCG signals from the wearable patch with finite impulse response (FIR) Kaiser window band-pass filters (cut-off frequencies: 10-30 Hz for the ECG and 1-40 Hz for the SCG signals). We chose these cut-off frequencies for the ECG signal to reduce the T-wave interference and amplify the R-wave for better R-peak detection in the subsequent signal processing step [31], [32]. We used the cut-off frequencies for the SCG signals to remove out-of-band noise without distorting the shape of the SCG signals [33]. After the filtering step, we computed a fourth SCG signal representing the accelerometer magnitude (SCG_{Mag}) using the three SCG axes already obtained (SCG_{HtoF}, SCG_{Lat}, SCG_{DV}) according to the following formula:

$$SCG_{\mathrm{Mag}} = \sqrt{(SCG_{\mathrm{HtoF}}{}^2 + SCG_{\mathrm{Lat}}{}^2 + SCG_{\mathrm{DV}}{}^2)} \qquad (1)$$

As further processing depends on high quality ECG signals, we used an automated artifact detection algorithm to identify motion artifacts in the ECG signal and excluded these corrupted segments from further analysis, as shown in Fig. 2(b). We initially segmented all the wearable signals into 30 second windows. For each window, we detected the upper (E_u) and lower (E_l) envelope of the data and computed a difference E_d (= E_u – E_l). We then calculated the mean (μ) and standard deviation (σ) of E_d throughout the 30 second frame. We deemed an ECG signal segment to contain artifacts when the E_d of that specific portion was greater or less than 3σ from the μ [34]. We removed frames containing artifact from all wearable signals using the timestamps for the affected segments. The artifact removal technique removed approximately 5% heartbeats in two subjects for the treadmill protocol and 5% heartbeats in one subject for the outside walking protocol. For other subjects, it removed less than 1% of the heartbeats.

2) Moving Ensemble Averaging

After removing sections with artifacts, we amplitude normalized the ECG (in the 30 second frame) and used the Pan Tompkins method [35], [36] to detect the R-peaks of the ECG signal. We segmented the wearable signals (ECG, four axes of SCG, AP signal) into individual heartbeats using the R-peaks of the ECG signal. We cropped each heartbeat to a duration of 500 ms after the R-peak. For heartbeats where the heart rate was greater than 120 beats per minute (and thus a second heartbeat appeared in the 500 ms window), we zero-padded the portion of the second heartbeat, to nullify the effect of the second heartbeat in the feature extraction process later. We chose the duration of 500 ms based on our previous experience with SCG signals and generic feature extraction processes [33], as most of the relevant systolic cardiac events of interests (e.g., aortic valve opening and closing) occur within this time frame from the corresponding R-peak of ECG. We chose a constant time window to crop the ECG, SCG, and AP signals to have a repeatable generic feature extraction process (explained in later section).

For all signals in each heartbeat, we averaged ten consecutive heartbeats in order to reduce noise and motion artifacts [37], obtaining ensemble averaged heartbeat frames across the whole recording. With a step-size of one heartbeat, this process resulted in a total of 64,349 ensemble averaged heartbeats from 17 subjects for the treadmill protocol and 37,485 ensemble averaged heartbeats from 15 subjects for the outside walk protocol. We calculated the R-to-R interval for each heartbeat and averaged this value across 10 heartbeats in the same way. We used R-to-R interval as a feature in subsequent modeling steps. We also calculated a mean AP signal value for each heartbeat and similarly averaged this value across 10 measurements. We used this value as an AP signal feature for modeling. Fig 2(c) shows an overview of the ensemble averaging technique for a single axis of SCG.

3) Feature Extraction from the SCG signals

Using an automated algorithm, we extracted 28 time domain, 17 amplitude and 17 frequency domain features (62 total) from the ensemble-averaged waveforms for each of the four SCG signals. The list of extracted features is provided in Table I. We extracted the amplitude and time domain features from the time-domain representation of the SCG signals, and the frequency domain features from the power spectral density (PSD) estimate of the SCG signals. Peaks and valleys in the averaged frames were ranked according to their amplitudes, and the highest and second highest amplitude were used. Location was calculated as the distance from the corresponding R-peak in ms. Width was calculated as the width of the peak or valley at half-prominence, in ms. For prominent frequencies in the PSD, peaks in the PSD of the frame were ranked according to their amplitudes, and the highest, second highest and third highest amplitude were used to locate the first, second and third prominent frequency accordingly. We evaluated different combinations of feature sets for performance in estimating VO_2, which is further explained in details in Section III.A.

4) Feature Extraction from the ECG Signal

We calculated the R-to-R interval and instantaneous heart rate for each heartbeat and averaged them with the same ensemble averaging method described in the "Moving Ensemble Averaging" sub-section. We used both of these as ECG signal features.

5) Feature Extraction from the AP signal

For each protocol, we selected AP signal values corresponding to the first 50 averaged heartbeats for each subject as baseline values for that subject. Averaging these, we obtained a single baseline AP value (AP_{BL}) for each subject. For all AP values for each subject, we obtained a delta pressure (δP) measurement using the following formula:

$$\delta P_{\mathrm{i}} = AP_{\mathrm{i}} - AP_{\mathrm{BL}} \quad (2)$$

where, i=1:number of averaged heartbeats for a particular subject. δP was the only feature from the AP signal used. With 62 features from each SCG signal, two features from ECG and one feature from AP, we had a total of 251 features per ensemble averaged heartbeat.

TABLE I
SCG FEATURES EXTRACTED

Signals	Feature Names	Number of Features		
		Ampl.	Time	Freq.
SCG (0-200 ms)	Highest and second highest peak (Ampl., Loc. and Width)	2	4	
	Lowest and second lowest valley (Ampl., Loc. And Width)	2	4	
	Number of peaks and valleys		2	
	First and second peak (Ampl., Loc. And Width)	2	4	
	First and second valley (Ampl., Loc. And Width)	2	4	
	Highest peak of absolute signal (Ampl., Loc. And Width)	1	2	
SCG (200-500 ms)	Highest peak (Ampl., Loc. and Width)	1	2	
	Lowest valley (Ampl., Loc. And Width)	1	2	
	Number of peaks and valleys		2	
	Highest peak of absolute signal (Ampl., Loc. And Width)	1	2	
SCG AUS	(0-100 ms), ..., (400-500 ms)	5		
SCG PSD Band Power	(0-3 Hz), (3-6 Hz), ..., (27-30 Hz) and (0-500 Hz)			11
SCG PSD	First, second and third prominent frequency (Ampl. and Freq.)			6
Single Axis Total		17	28	17
4-Axes Total		68	112	68

Ampl: amplitude, Freq: frequency, Loc: location, AUS: area under signal, PSD: power spectral density.

6) Parameter Estimation from the BxB Data

First, we smoothed the time-synchronized BxB data with a 5 point moving average window to reduce noise [38]. We computed averaged VO_2 measurements corresponding to each ensemble averaged waveform with the time-stamps taken during recording. These VO_2 values functioned as the target variable in the regressor for each ensemble averaged heartbeat. We used Matlab® to conduct all signal processing and feature extraction steps.

D. Machine Learning and Regression

1) Regression Model

In previous work, researchers used linear regression [39], multiple linear regression [30] and non-linear regression [19] models to estimate VO_2 and/or EE. In our prior work with pre-ejection period (PEP) estimation from SCG [33], we have demonstrated that non-linear ensemble methods outperform linear regression models in developing global regression models for PEP estimation from wearable SCG sensors. For that reason, we have explored different linear and non-linear ensemble regression models in our initial analysis to estimate VO_2 from wearable signals. From our initial analysis, we chose ensemble regression models over linear regression models.

Before training a regression model to estimate VO_2, we removed outlying heartbeats from the ensemble averaged SCG waveforms. For each subject, we designated the first 50 averaged heartbeat frames from the signals measured at rest as baseline heartbeat frames and averaged all the features (for a particular feature set consisting of f features from SCG, ECG and AP signals) of the 50 frames to create a baseline feature distribution. We calculated the Mahalanobis distance [40] between the baseline feature distribution and each averaged heartbeat frame for a particular subject. Our underlying hypothesis was that the wearable signals would change morphologically with various intensities of exercise. We expected greatest morphological changes compared to baseline at peak exercise and expected the Mahalanobis distance to capture this variation. We deemed a particular heartbeat frame an outlier based on the interquartile range criteria described in [34] and excluded these frames from the dataset. One thing to note is that the previous outlier removal step for the ECG signal using an envelope-based detection technique was used to remove artifact-affected ECG segments, which did not involve SCG artifacts in the decision-making process, but rather focuses on the ECG alone. Nevertheless, some of the SCG heartbeats could also be heavily affected by motion artifacts and thus a second outlier removal step

was required based on SCG characteristics. We used the Mahalanobis distance calculated for each frame as a feature in the regression model, which makes the total number of features equal to *f+1*. This outlier heartbeat removal technique removed approximately 5% heartbeats for the outside walking protocol and 10% heartbeats for the treadmill protocol.

After removing outlier heartbeats, we trained different regression models to estimate the target variables from the features extracted in the Method section. From our initial analysis, Extreme Gradient Boosting (XGBoost) [41] regression outperformed other regression methods. For that reason, we chose XGBoost regression for detailed analysis in this work. XGBoost is a decision-tree based ensemble algorithm that uses a gradient boosting [42] framework. It is an example of an ensemble method [43] that is computationally efficient, parallelizable, able to handle missing values, and able to be pruned/regularized to avoid overfitting. We fit an XGBoost regressor on the extracted features for all ensemble averaged heartbeats to estimate corresponding target VO_2 values. We then used this model to estimate VO_2 values for unseen heartbeat frames as represented by the same feature sets. We performed this process with different combinations of feature sets extracted from the SCG, ECG, and AP signals, optimizing hyper-parameters with a grid search. The final hyper-parameters are as follows: learning rate=0.05, max_depth=10, subsample=0.6, colsample_bytree=0.7, n_estimators=100, min_child_weight=2, gamma=0.3. We used Python 3.6 for all machine learning techniques.

2) Cross-validation and Model Evaluation

We used leave-one-subject-out (LOSO) cross-validation for *n* subjects for both protocols. At each fold we trained an XGBoost regressor on the data from *n-1* subjects, leaving one subject out. We then predicted VO_2 for the left-out subject, repeating this *n-1* more times with a different subject excluded each time. As a result, we obtained predictions for all ensemble averaged heartbeats. We calculated a root mean squared error (RMSE) between the estimated target variable ($VO_{2,e}$) and the ground truth target variable acquired from the BxB data ($VO_{2,a}$):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(VO_{2,e\ (i)} - VO_{2,a\ (i)})^2} \qquad (3)$$

where N is the number of ensemble average heartbeats for a particular subject. We calculated the cross-validated RMSE as the average of the RMSE scores from n folds. In this way, we trained two different global regression models for two exercise protocols. We also calculated coefficients of determination (R^2) between the true values and the cross validated predictions of VO_2 across all subjects.

3) Regression for Different Feature Sets

Repeating this regression and cross-validation approach for 15 different combinations of feature sets (shown in Table II), we compared the resulting cross-validated RMSE scores to see which performed better in estimating VO_2, using an XGBoost regressor. We performed statistical analysis on the cross-validation results from the different feature sets.

In literature, researchers have used HR to estimate VO_2 using simple correlation analysis and linear regression with/without cross-validation [19], [44]–[46]. To compare the performance of our proposed algorithm with HR-based models, we trained a simple linear regression model with LOSO cross-validation to estimate VO_2 from HR only for the two protocols separately and calculated the RMSE and R^2 (also reported in Table II). We later compared the results obtained from this HR-based linear regression model with our proposed algorithm (using XGBoost). Additionally, we performed simple correlation analysis (without any cross-validation) between instantaneous HR and VO_2 for the two protocols and calculated subject-wise R^2 and global R^2 (with all data for a particular protocol) to compare our proposed algorithm with a simple HR-based model.

4) Regression with Feature Selection

We performed different feature selection techniques to select K (=25) features out of 251 (SCG, ECG, and AP) features for the estimation of VO_2, to explore how a subset of selected features can perform compared to different feature sets described in the section above. Based on our initial analysis, we chose the feature selection technique using sequential forward selection (SFS) [47] for detailed analysis. We performed the SFS method on the whole dataset separately for the two protocols and used those selected features to train an XGBoost regression model per protocol. We calculated the RMSE and R^2 following the same LOSO cross-validation approach described above and compared the results to the best-performing feature set for the two protocols respectively.

5) Cross-Evaluation of Regression Models

For cross-evaluation of the two models generated from the two protocols, we trained an XGBoost regressor on all the data from the treadmill protocol and predicted VO_2 for the data from the outside protocol and vice versa. We calculated subject-wise RMSE and overall R^2 (as described before) to compare the estimation accuracy and generalizability of the two models. We used only the best-performing feature set from the combination of SCG and ECG features to implement this cross-evaluation.

E. Comparison with Similar Studies

We compared the performance of our proposed algorithm to the performance of similar studies [10], [21]. Besides RMSE and R^2 for instantaneous VO_2 estimation, we calculated percentage estimation error for VO_2 consumed over one minute intervals using the equation below, following the works reported in [10], [21].

$$error = \frac{|\sum VO_{2,e} - \sum VO_{2,a}|}{\sum VO_{2,a}} \times 100 \qquad (4)$$

For each subject, we calculated subject-wise mean of the percentage errors and calculated the median percentage error (MPE) following [10] for all subjects. We calculated the MPE using the best performing feature sets for the two protocols.

F. Statistical Analysis

We performed statistical analysis on cross-validated RMSE results of different feature sets. We used multiple comparison tests on the RMSE results from the *n* cross validation folds. We

employed the Friedman test to detect if statistical differences existed between our results and the Wilcoxon signed rank test for post-hoc testing. Additionally, for the post-hoc testing, we applied the Benjamini-Hochberg correction for multiple comparisons on the p-values. To explore the importance of the single AP feature, we performed Wilcoxon signed rank test on the subject-wise RMSE results with and without the AP feature. Details on these statistical tests and the reasons behind their use are discussed in [48]. In this work, we considered p-values below 0.05 statistically significant.

III. Results and Discussion

Fig. 3(a) shows the wearable signals (filtered ECG, filtered SCG_{DV} and raw AP signals) and the VO_2 target variable for both protocols for one subject. Fig 3(b) and (c) show the averaged SCG_{DV} waveform at different exercise intensity levels for treadmill and outside walking exercise respectively. SCG_{DV} waveforms exhibit changes in amplitude with incremental changes in exercise during treadmill walking as shown in Fig. 3(a) (left) and (b). VO_2 also exhibits incremental changes during treadmill exercise as expected in Fig. 3(a). AP signals seem somewhat stable in general during treadmill exercise. In the case of outdoor walking in Fig. 3(a) (right), AP signals track the altitude of the terrain well. As expected, VO_2 changes with the gradation. In contrast, changes in SCG_{DV} waveforms shown in Fig. 3(a) (right) and (c) are not very apparent.

Overall, Fig. 3 shows that the SCG signal amplitude changes significantly with the intensity of exercise, which results from physiological changes as well as motion artifacts. However, we believe movement artifacts were not synchronized to the heartbeat timings and the impacts of motion artifacts on the SCG heartbeats (and extracted features from them) were reduced accordingly using a moving ensemble average of 10 heartbeats. Future analysis should look into component analysis of the signals to isolate the relevant cardiac and motion information from the signals. As the SCG signal varies among subjects, we refrained from putting any emphasis on the signal shape and peaks to correlate with underlying cardiac events [33]. Rather we wanted to explore how generic time, frequency and amplitude features extracted from SCG at various exercise/movement intensities can be used to develop a global regression model to estimate instantaneous VO_2 in both controlled and uncontrolled settings. This reduces the complexity of the signal processing by avoiding the inter-subject variability and enables repeatable feature extraction and estimation of VO_2 in a population-level model.

Upon inspection of Fig. 3, it is apparent that the addition of cardio-mechanical (i.e. SCG) and environmental signals (i.e. AP) to HR-based models has the potential to improve estimations of instantaneous VO_2 during both controlled and uncontrolled activity. The details of estimation results are given in the following sections.

A. Comparison of Different Feature Sets of SCG with ECG

Table II shows the RMSE in ml/kg/min and R^2 values for different combinations of feature sets extracted from the wearable signals. Statistically significant differences existed in these results according to the Friedman test ($p<0.05$). We performed Wilcoxon signed rank tests on the different feature sets to investigate the significance of their differing accuracy values. All the feature sets described in this table included an AP signal feature (except HR-based linear regression model) in addition to the features explicitly stated.

TABLE II
RMSE (*ml/kg/min*) and R^2 for VO_2 estimation from different feature sets of SCG (Amplitude, *Ampl*, Frequency, *Freq* and Time, *Time*) and ECG using XGBoost, and HR using linear regression model

	Treadmill Protocol		Outside Walking Protocol	
Feature Set	RMSE	R^2	RMSE	R^2
Ampl	4.06±1.06	0.76	4.8±1.53	0.52
Freq	**3.68±0.98**	0.77	4.85±1.31	0.57
Time	5.42±1.39	0.45	5.13±1.18	0.45
ECG	7.48±1.83	0.17	5.81±1.07	0.4
Ampl+ECG	4.24±1.18	0.72	4.46±1.56	0.58
Freq+ECG	3.99±1.28	0.71	**4.30±1.47**	**0.64**
Time+ECG	5.07±1.79	0.5	4.89±1.66	0.47
Ampl+Freq	3.78±0.98	**0.78**	4.79±1.53	0.54
Ampl+Time	4.04±1.27	0.68	4.93±1.62	0.47
Freq+Time	3.9±1.33	0.67	4.92±1.41	0.48
Ampl+Freq+ECG	3.98±1.27	0.75	4.52±1.52	0.59
Ampl+Time+ECG	4.27±1.54	0.64	4.81±1.67	0.48
Freq+Time+ECG	4.27±1.61	0.61	4.69±1.65	0.51
Ampl+Freq+Time	3.87±1.18	0.73	4.95±1.57	0.47
Ampl+Freq+Time+ECG	4.12±1.4	0.68	4.75±1.66	0.5
HR[a]	6.31±1.72	0.44	5.94±1.76	0.35
Selected 25 Features[b]	3.76±1.15	0.77	4.28±1.44	0.63

[a]A simple linear regression model was used for HR only. For other feature sets, XGBoost regression model was used to generate the reported results.
[b]Selected using SFS method

As shown in table II, of the single SCG feature sets, frequency domain features achieved the lowest RMSE and highest R^2 for the treadmill protocol. Amplitude features were slightly worse ($p>.05$) and time domain features performed the poorest ($p<.05$ compared to both frequency and amplitude). For the outdoor protocol, frequency features achieved the best R^2 and had an RMSE only slightly above ($p>.05$) that of amplitude features (with a narrower confidence interval). Time domain features once again performed the worst, though not significant ($p>.05$). From these results it appears that frequency domain features provided the most salient information for estimating VO_2 from SCG in both settings.

Better performance of frequency domain features in the estimation of VO_2 is understandable as exercise leads to substantial changes in the shape and timing of waveforms. For example, the shortening of isovolumetric contraction time associated with increased sympathetic tone compresses the SCG waveform in time and thus increases high frequency components [29]. VO_2 relates to Stroke Volume [49] which has been shown to have a relation with the amplitude features of the SCG signal [25]. This result is consistent with [28], [50], where researchers have used frequency domain features of the SCG signal to assess clinical state for patients with HF. When comparing frequency and amplitude features, frequency features performed slightly ($p>0.05$) better. Our results show that exercise induced changes of VO_2 not only change the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.3009903, IEEE Journal of Biomedical and Health Informatics

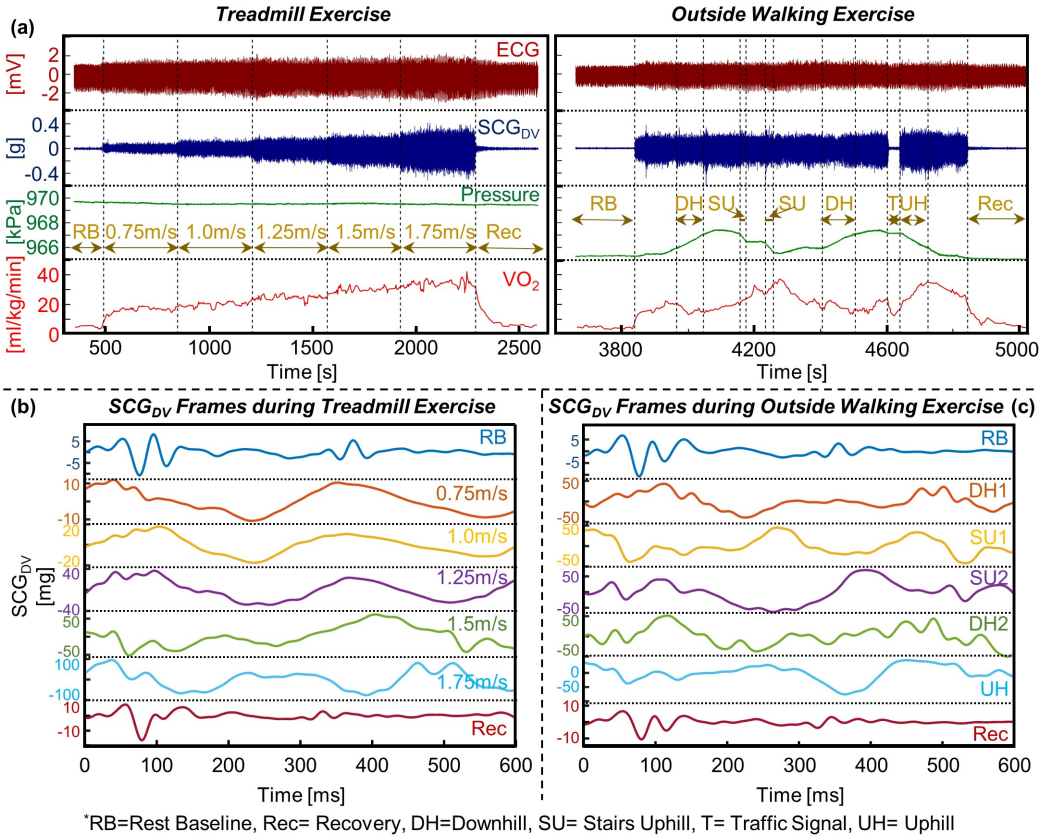

Fig. 3. (a) Illustration of wearable signals (ECG, SCG and AP) with corresponding VO_2 from COSMED K5 system at different exercise intensity level for both **(left)** treadmill exercise and **(right)** outside walking exercise. (b) Ensemble averaged heartbeats of SCG_{DV} signal at different exercise intensity levels during treadmill exercise. (c) Ensemble averaged heartbeats of SCG_{DV} signal at different exercise intensity levels during outside walking exercise.

amplitude of the SCG signals but also the signal power at different frequency bands. Frequency domain features captured these changes better than amplitude features.

ECG features alone performed worse than all three SCG features alone in both protocols ($p<0.05$ compared to all three feature sets of SCG for the treadmill protocol and $p<0.05$ compared to amplitude and frequency feature sets of SCG for the outside walking protocol), using XGBoost regression algorithm. As other studies have demonstrated high linear correlations between HR and VO_2 [19], [44]–[46], the comparatively poor performance of ECG features (instantaneous HR and R-to-R interval) in our approach is likely attributable to the overly-complex nature of an XGBoost regression model and/or the addition of the R-to-R interval feature. To compare our results with the common HR-based approach, we also fit a simple linear regression model with HR only to estimate instantaneous VO_2 using the same LOSO cross-validation approach, which achieved an RMSE of 6.31±1.72 and R^2 of 0.44 for the treadmill protocol and an RMSE of 5.94±1.76 and R^2 of 0.35 for the outside walking protocol. Still these results are significantly poorer ($p<0.05$) compared to the amplitude and frequency domain features of SCG. Separate from the HR-based simple linear regression model with LOSO cross-validation, we also performed a simple correlation analysis between instantaneous HR and VO_2 across all subjects, which resulted in an overall R^2 of 0.49 and 0.42 for treadmill and outside walking protocol respectively. A similar analysis on each subject individually resulted in a higher R^2 of 0.73±0.11 and 0.71±0.16 for treadmill and outside walking protocol respectively. The lower value of the global R^2 compared to the subject-wise R^2 is in agreement with existing literature [45], [46]. For this reason, researchers often use %VO_2-max and %HR-max when attempting to show population-level relationships between VO_2 and HR rather than their raw values directly [45], [46], [51]–[53]. Overall, these results show the benefit of incorporating cardio-mechanical information from SCG into a complex machine learning algorithm for the development of a global regression model to estimate instantaneous VO_2 compared to simple linear models involving only HR-based information.

When combining different feature sets of SCG with ECG in XGBoost regressors, we achieved our best results (i.e. lowest RMSE and highest R^2) on the treadmill protocol using amplitude and frequency features of SCG (RMSE of 3.78±0.98 ml/kg/min and R^2 of 0.78), which is significantly lower ($p<0.05$) than amplitude features alone. Still the lowest RMSE for the treadmill protocol was obtained using frequency domain features alone (RMSE of 3.68±0.98 ml/kg/min with frequency domain features alone vs. 3.78±0.98 with amplitude and frequency domain features together, $p>0.05$). For the outdoor protocol, we obtained our best results using frequency features of SCG and ECG features (RMSE of 4.30±1.47

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.3009903, IEEE Journal of Biomedical and Health Informatics

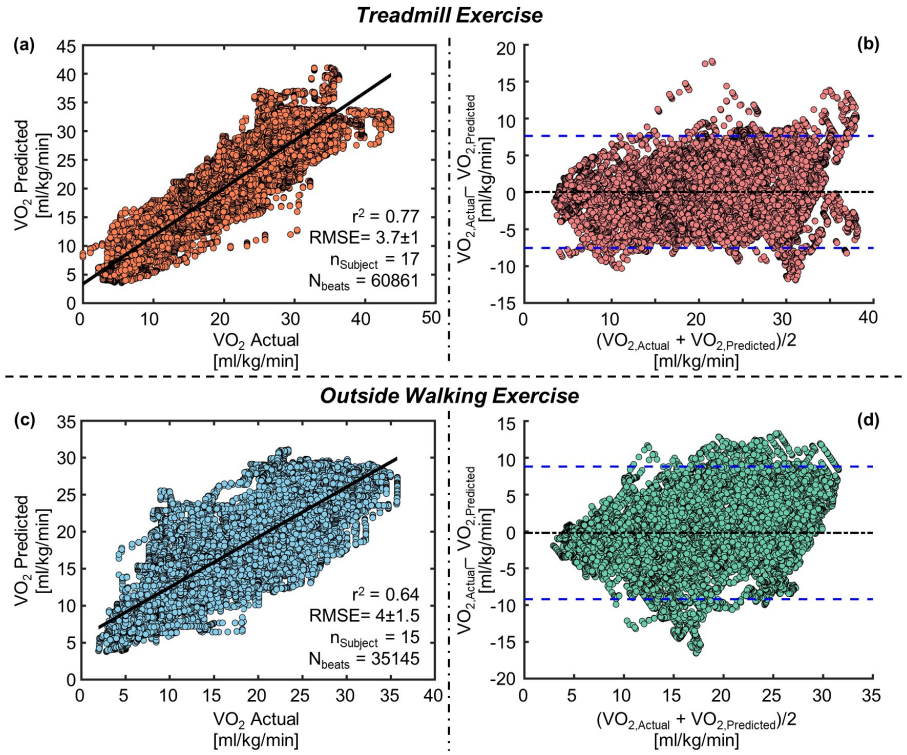

Fig. 4. (a) Correlation analysis for VO_2 predicted vs VO_2 actual (b) Bland-Altman analysis for VO_2 predicted and VO_2 actual for treadmill exercise. (c) Correlation analysis for VO_2 predicted vs VO_2 actual (d) Bland-Altman analysis for VO_2 predicted and VO_2 actual for outside walking exercise. In the Bland-Altman plots, the solid black line indicates the mean while the blue dashed lines indicate mean $\pm$ 1.96 x standard deviation.

ml/kg/min and R^2 of 0.64). These results were significantly better ($p<0.05$) than those from frequency, amplitude, time and ECG features alone. Best results in RMSE and R^2 values for each protocol are shown in bold in Table II. Fig. 4 provides a correlation analysis and Bland-Altman analysis of actual VO_2 values and estimated VO_2 values, using the feature set with the lowest RMSE for each protocol. It is apparent that regression models can generally estimate instantaneous VO_2 well on a heartbeat by heartbeat basis.

Fig. 5 shows examples of best and worst estimations of VO_2 compared to actual VO_2 for both protocols. The same model used in Fig. 4 generated the VO_2 estimations here. Even when achieving the worst results in both protocols as shown in Fig. 5 (b) and (d), the algorithm still tracks relative changes well despite overestimating overall VO_2 values. Hence, we see a consistent offset between actual and predicted values in both cases.

From Table II in the case of the treadmill protocol, adding ECG features independently to both the amplitude feature set and the frequency feature set increased the error in both cases. This is as expected because performance of ECG features was the worst among all the feature sets for the treadmill protocol. For the outside walking protocol, adding ECG features reduced the RMSE for these same two feature sets ($p<0.05$). For better VO_2 estimation, the selection of feature sets for a global model should incorporate domain knowledge of cardio-electromechanical responses to ranges of exercise and activity.

Overall from Table II, the estimation results were better for the treadmill protocol than for the outside walking protocol. This is expected as the treadmill protocol took place indoors with standardized speeds and conditions whereas the outside walking protocol was completed at the subject's pace with variable atmospheric conditions depending on the day. Future studies should examine wider varieties of exercise with subjects of broader age range and health status to apply this methodology in estimating instantaneous VO_2 throughout daily activities.

B. Quantification of Additional Benefit from AP Signal

To assess the addition of environmental features in estimating VO_2, we compared results from the best performing feature set for both protocols with and without features from the AP signal. We conducted statistical analysis using a Wilcoxon signed rank test on the paired RMSE values with and without features from AP signal. For the treadmill protocol, there were no significant changes ($p>0.05$) in RMSE from adding features from AP signal. In the case of the outside protocol, adding features from the AP signal significantly reduced ($p<0.05$) the RMSE from 4.72 ± 1.59 to 4.30 ± 1.40 ml/kg/min. This result is as expected from Fig. 3(a), where it can be clearly seen that the AP signal captures valuable information regarding altitude during the outside walk. This result is consistent with the work of Lester et al. [21] when they showed that incorporating grade from barometric pressure and/or GPS improved their accuracy in estimating caloric expenditure in outside settings. Thus, the combination of cardiac electro-mechanical parameters and environmental (AP) context, presented for the first time in this paper, outperforms physiological measurements alone in estimating VO_2.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.3009903, IEEE Journal of Biomedical and Health Informatics

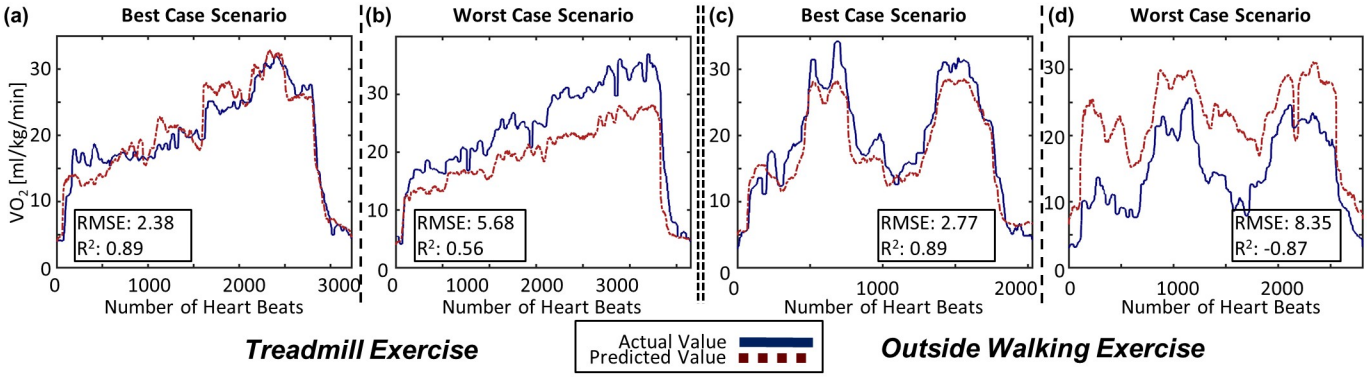

Fig. 5. Example of subject-wise VO_2 prediction for both exercise tasks: (a) Best case scenario and (b) worst case scenario for treadmill exercise. (c) Best case scenario and (d) worst case scenario for outside walking exercise.

C. Regression with Feature Selection

The list of the selected features using the SFS algorithm is provided in the appendix section. Of the 25 selected features for treadmill protocol, one feature was ECG, 11 were SCG amplitude, 6 were SCG frequency, and 7 were SCG time-domain features (10 were from SCG_{HtoF}, 4 were from SCG_{Lat}, one was from SCG_{DV}, and 9 were from SCG_{Mag}). In the case of the outside walking protocol, one feature was ECG, one was AP, 8 were SCG amplitude, 10 were SCG frequency, and 5 were SCG time-domain features (10 features were from SCG_{HtoF}, 4 were from SCG_{Lat}, 4 were from SCG_{DV}, and 5 were from SCG_{Mag}). Comparatively higher number of SCG_{HtoF} features in the selected features are in accordance with our previous experience in PEP estimation from the SCG features [33], [54].

The regression models with the top 25 features achieved an RMSE of 3.76±1.15 ml/kg/min and R^2 of 0.77 for the treadmill exercise and an RMSE of 4.28±1.44 ml/kg/min and R^2 of 0.63 for the outside walking exercise. These results are very similar ($p>0.05$) with the ones we obtained using the best-performing feature sets for the two protocols respectively. Overall, these results show the importance of different amplitude, frequency, and time-domain features of SCG as well as the importance of features from different SCG axes over the typical use of features from SCG_{DV} only. Future work should verify these initial findings from this pilot study in a larger, more diverse population.

D. Cross-Evaluation of Regression Models

The model trained on all the data from the treadmill protocol and tested on the data from the outside protocol resulted in an RMSE of 5.8±1.62 and R^2 of 0.49, whereas, the model trained on all the data from the outside walking protocol and tested on the data from the treadmill protocol resulted in an RMSE of 5.26±1.92 and R^2 of 0.66. The higher accuracy with the model trained using data from the outside walking protocol compared to the model trained using the data from the treadmill protocol is expected, as the outside protocol includes a variety of activities (rest, uphill/downhill walking, stairs and recovery) whereas the treadmill protocol only covers incremental exercise. The model trained on treadmill data was not exposed to the variety of activities and their corresponding physiological changes (which resulted in the changes in SCG and ECG signals) that it encountered during testing on the outside walking protocol and thus could not estimate instantaneous VO_2 as well in these unseen conditions. This cross-evaluation was performed using the combination of frequency features of SCG and ECG. Overall, this cross-evaluation showcases the generalizability of our approach and demonstrates the necessity of training models on data collected in uncontrolled settings to estimate instantaneous VO_2 in daily life environments. These results suggest that training models on data from uncontrolled settings has the potential to additionally improve performance for specific exercise applications, yet data from controlled laboratory environments do not suffice for deploying the system ubiquitously.

E. Comparison with Similar Studies

Our MPE when estimating VO_2 over one minute intervals was 15.8% for the treadmill protocol and 20.5% for the outside protocol. Compared to results from Shcherbina et al. [10] on commercial devices in controlled laboratory settings, these results are good (MPE of 15.8% in our case compared to MPE of 27.4% in [10]). Similarly, if converted to percentage accuracy (100%-MPE), they are comparable with the work of Lester et al. [21] for both protocols (our work: 84.2% and 79.5% for controlled and uncontrolled settings respectively, Lester et al.: 89.52% and 79.8% respectively). However, it is important to note that these studies and ours all differ in their activity protocols, subject populations, and output granularities (instantaneous readings vs averages over intervals, output for entire protocols vs only select segments etc.) and are thus not directly comparable. In our case, we have used the data from the whole protocol and estimated instantaneous VO_2 for the full duration. We report our error for both instantaneous estimation as well as over one minute intervals. As mentioned earlier, our study involves a population of 17 relatively homogeneous subjects for the treadmill protocol and 15 for the outdoor protocol.

IV. Conclusion and Future Work

In this paper, we have estimated instantaneous VO_2 during controlled treadmill and uncontrolled outside walking exercise using features from SCG, ECG and AP signals captured with a wearable patch placed on a subject's mid-sternum. We have developed a global regression model for VO_2 estimation using state-of-the-art machine learning algorithms validated with

leave-one-subject-out cross-validation. We have demonstrated that adding cardio-mechanical and environmental sensing modalities to traditional HR-based models improves estimation accuracy significantly. Regarding SCG features, frequency domain features provided the most valuable information for VO_2 estimation. Overall, this work demonstrates the capability of an unobtrusive wearable patch to track VO_2 and EE continuously in daily life and exercise activities. Success in this regard represents a considerable step towards the vision of personalized, ubiquitous monitoring of fitness levels, weight-loss programs, and lifestyle habits in healthy individuals. It additionally suggests potential applications to the tracking of disease progression and recovery progress for clinical purposes. The low-cost nature of our system could make these benefits accessible to the full general public.

As we collected data from healthy young subjects only, future studies should increase the number of subjects and broaden diversity in age and exercise ability. Additionally, we collected data during treadmill and outside walking exercise only for this proof of concept study. Future work should investigate the use of the sensing system and VO_2 estimation algorithm for different and more complex daily life activities, which will potentially enable the use of the device in ubiquitous monitoring of EE. We extracted generic time, frequency and amplitude features of the SCG signals, and have not considered how the features can be affected by motion and exercise intensities. Future work should investigate the effect of the number of steps, step length, stride time, and velocity on the SCG signals, and corresponding features. We removed portions of the signals corrupted with motion artifact with an outlier detection algorithm. Future studies should approach the reduction of motion artifact from both a hardware and signal processing perspective to potentially improve estimation. Our only environmental feature in this study was measurement of pressure changes. Future studies should evaluate the potential benefits of incorporating temperature, humidity and other environmental features in the global model, as well as instantaneous velocity and respiration-related features. We explored the relationship between SCG features and VO_2. Future studies should investigate the mechanisms that underlie these relationships. Finally, this study demonstrated an approach for measuring VO_2 with a wearable device at sub-maximal effort levels. Future studies should investigate how a similar approach might function at all levels of exertion to potentially enable estimation of VO_2 max and CRF.

APPENDIX

A. Selected Wearable Features using SFS

For treadmill protocol: R-to-R interval, amplitude of first peak, valley and absolute maxima (0-200 ms) of SCG_{HtoF}, amplitude of second maxima, second peak and first valley (0-200 ms) of SCG_{Mag}, AUS (0-100 ms) of SCG_{DV}, AUS (0-100, 200-300, 300-400, 400-500 ms) of SCG_{Mag}, SCG_{Lat} band Power (15-18 Hz), SCG_{HtoF} band Power (12-15 Hz), most prominent and second prominent frequency in PSD of SCG_{HtoF}, amplitude of second prominent frequency in PSD of SCG_{HtoF} and SCG_{Mag}, location of second minima and first valley (0-200 ms) of SCG_{Lat}, width of first absolute maxima (200-500 ms) of SCG_{Lat}, width of second minima, location and width of second peak (0-200 ms) of SCG_{HtoF}, and location of first absolute maxima (0-200 ms) of SCG_{Mag}.

For outside walking protocol: R-to-R interval, delta pressure, amplitude of first maxima, minima and absolute maxima (200-500 ms) of SCG_{HtoF}, amplitude of first maxima, absolute first maxima, first peak (0-200 ms) of SCG_{Mag}, AUS (200-300 ms) of SCG_{HtoF} and SCG_{DV}, SCG_{Lat} band power (12-15 Hz), SCG_{HtoF} band power (15-18, 18-21, 21-24, 27-30 Hz), SCG_{DV} band power (21-24 Hz), SCG_{Mag} band power (24-27 Hz), most prominent frequency in PSD of SCG_{HtoF}, amplitude of second prominent frequency in PSD of SCG_{Lat} and SCG_{DV}, location of first maxima and width of second minima (0-200 ms) of SCG_{Lat}, width of first maxima (200-500 ms) of SCG_{HtoF}, number of valleys (0-200 ms) of SCG_{DV}, and location of first minima 200-500 ms of SCG_{Mag}.

DISCLOSURE

Omer T. Inan is a Scentific Advisor to Physiowave, Inc., a manufacturer of ballistocardiogram measurement weighing scales.

REFERENCES

[1] R. Malhotra, K. Bakken, E. D'Elia, and G. D. Lewis, "Cardiopulmonary exercise testing in heart failure," *JACC Heart Fail*, vol. 4, no. 8, pp. 607–16, 2016.

[2] J. Myers, R. Arena, L. P. Cahalin, V. Labate, and M. Guazzi, "Cardiopulmonary exercise testing in heart failure," *Curr Probl Cardiol*, vol. 40, no. 8, pp. 322–72, 2015.

[3] M. J. Haykowsky, C. R. Tomczak, J. M. Scott, D. I. Paterson, and D. W. Kitzman, "Determinants of exercise intolerance in patients with heart failure and reduced or preserved ejection fraction," *J Appl Physiol (1985)*, vol. 119, no. 6, pp. 739–44, 2015.

[4] A. A. Ganju, A. B. Fuladi, B. O. Tayade, and N. A. Ganju, "Cardiopulmonary exercise testing in evaluation of patients of chronic obstructive pulmonary disease," *Indian J Chest Dis Allied Sci*, vol. 53, no. 2, pp. 87–91, 2011.

[5] M. T. Hamilton, D. G. Hamilton, and T. W. Zderic, "Role of low energy expenditure and sitting in obesity, metabolic syndrome, type 2 diabetes, and cardiovascular disease," *Diabetes*, vol. 56, no. 11, pp. 2655–67, 2007.

[6] A. M. Fair and K. Montgomery, "Energy balance, physical activity, and cancer risk," *Methods Mol Biol*, vol. 472, pp. 57–88, 2009.

[7] L. Guidetti, M. Meucci, F. Bolletta, G. P. Emerenziani, M. C. Gallotta, and C. Baldari, "Validity, reliability and minimum detectable change of cosmed k5 portable gas exchange system in breath-by-breath mode," *PLoS One*, vol. 13, no. 12, p. e0209925, 2018.

[8] I. Perez-Suarez, M. Martin-Rincon, J. J. Gonzalez-Henriquez, C. Fezzardi, S. Perez-Regalado, V. Galvan-Alvarez, J. W. Juan-Habib, D. Morales-Alamo, and J. A. L. Calbet, "Accuracy and precision of the cosmed k5 portable analyser," *Front Physiol*, vol. 9, p. 1764, 2018.

[9] A. J. Cook, S. J. Redmond, G. D. Gargiulo, and T. J. Hamilton, "Techniques for measuring energy expenditure with portable devices," *2013 Ieee Tencon Spring Conference*, pp. 39–42, 2013.

[10] A. Shcherbina, C. M. Mattsson, D. Waggott, H. Salisbury, J. Christle, T. Hastie, M. T. Wheeler, and E. A. Ashley, "Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort," *Journal of Personalized Medicine*, vol. 7, no. 2, 2017.

[11] C. Dondzila and D. Garner, "Comparative accuracy of fitness tracking modalities in quantifying energy expenditure," *J Med Eng Technol*, vol. 40, no. 6, pp. 325–9, 2016.

[12] M. P. Wallen, S. R. Gomersall, S. E. Keating, U. Wisloff, and J. S. Coombes, "Accuracy of heart rate watches: Implications for weight management," *PLoS One*, vol. 11, no. 5, p. e0154420, 2016.

[13] A. W. Gorny, S. J. Liew, C. S. Tan, and F. Muller-Riemenschneider, "Fitbit charge hr wireless heart rate monitor: Validation study conducted under free-living conditions," *JMIR Mhealth Uhealth*, vol. 5, no. 10, p. e157, 2017.

[14] J. E. Sasaki, A. Hickey, M. Mavilia, J. Tedesco, D. John, S. Kozey Keadle, and P. S. Freedson, "Validation of the fitbit wireless activity tracker

for prediction of energy expenditure," *J Phys Act Health*, vol. 12, no. 2, pp. 149–54, 2015.

[15] T. C. Wong, J. G. Webster, H. J. Montoye, and R. Washburn, "Portable accelerometer device for measuring human energy-expenditure," *Ieee Transactions on Biomedical Engineering*, vol. 28, no. 6, pp. 467–471, 1981.

[16] C. V. Bouten, K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen, "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *IEEE Trans Biomed Eng*, vol. 44, no. 3, pp. 136–47, 1997.

[17] M. Miyatake, N. Nakamura, T. Nagata, A. Yuuki, H. Yomo, T. Kawabata, and S. Hara, "Vo2 estimation using 6-axis motion sensing data," *2016 10th International Symposium on Medical Information and Communication Technology (Ismict)*, 2016.

[18] P. Alinia, R. Saeedi, B. Mortazavi, A. Rokni, and H. Ghasemzadeh, "Impact of sensor misplacement on estimating metabolic equivalent of task with wearables," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 1–6, June 2015.

[19] A. J. Cook, B. Ng, G. D. Gargiulo, D. Hindmarsh, M. Pitney, T. Lehmann, and T. J. Hamilton, "Instantaneous vo2 from a wearable device," *Medical Engineering Physics*, vol. 52, pp. 41–48, 2018.

[20] S. P. Liu, R. X. Gao, D. John, J. W. Staudenmayer, and P. S. Freedson, "Multisensor data fusion for physical activity assessment," *Ieee Transactions on Biomedical Engineering*, vol. 59, no. 3, pp. 687–696, 2012.

[21] J. Lester, C. Hartung, L. Pina, R. Libby, G. Borriello, and G. Duncan, "Validated caloric expenditure estimation using a single body-worn sensor," *Ubicomp'09: Proceedings of the 11th Acm International Conference on Ubiquitous Computing*, pp. 225–234, 2009.

[22] L. Bouarfa, L. Atallah, R. M. Kwasnicki, C. Pettitt, G. Frost, and G. Z. Yang, "Predicting free-living energy expenditure using a miniaturized ear-worn sensor: An evaluation against doubly labeled water (vol 61, pg 566, 2014)," *Ieee Transactions on Biomedical Engineering*, vol. 61, no. 11, pp. 2818–2818, 2014.

[23] S. Brage, K. Westgate, P. W. Franks, O. Stegle, A. Wright, U. Ekelund, and N. J. Wareham, "Estimation of free-living energy expenditure by heart rate and movement sensing: A doubly-labelled water study," *Plos One*, vol. 10, no. 9, 2015.

[24] K. R. Westerterp, "Doubly labelled water assessment of energy expenditure: principle, practice, and promise," *European journal of applied physiology*, vol. 117, no. 7, pp. 1277–1285, 2017.

[25] O. T. Inan, P. F. Migeotte, K. S. Park, M. Etemadi, K. Tavakolian, R. Casanella, J. Zanetti, J. Tank, I. Funtova, G. K. Prisk, and M. Di Rienzo, "Ballistocardiography and seismocardiography: a review of recent advances," *IEEE J Biomed Health Inform*, vol. 19, no. 4, pp. 1414–27, 2015.

[26] M. M. H. Shandhi, J. Fan, A. J. Heller, M. Etemadi, O. T. Inan, and L. Klein, "Seismocardiography can assess cardiopulmonary exercise test parameters in patients with heart failure," *Journal of Cardiac Failure*, vol. 24, no. 8, pp. S124–S125, 2018.

[27] M. M. H. Shandhi, J. Fan, J. A. Heller, M. Etemadi, O. T. Inan, and L. Klein, "Seismocardiography and machine learning algorithms to assess clinical status of patients with heart failure in cardiopulmonary exercise testing," *Journal of Cardiac Failure*, vol. 25, no. 8, pp. S64–S65, 2019.

[28] M. M. H. Shandhi, S. Hersek, J. Fan, E. Sander, T. De Marco, J. A. Heller, M. Etemadi, L. Klein, and O. T. Inan, "Wearable patch based estimation of oxygen uptake and assessment of clinical status during cardiopulmonary exercise testing in patients with heart failure," *Journal of Cardiac Failure*, 2020.

[29] M. Etemadi, O. T. Inan, J. A. Heller, S. Hersek, L. Klein, and S. Roy, "A wearable patch to enable long-term monitoring of environmental, activity and hemodynamics variables," *IEEE transactions on biomedical circuits and systems*, vol. 10, no. 2, pp. 280–288, 2015.

[30] D. H. Kim, J. S. Cho, H. S. Oh, Y. J. Chee, and I. Y. Kim, "The estimation method of physical activity energy expenditure considering heart rate variability," *Cinc: 2009 36th Annual Computers in Cardiology Conference*, vol. 36, pp. 413–+, 2009.

[31] L. G. Tereshchenko and M. E. Josephson, "Frequency content and characteristics of ventricular conduction," *Journal of electrocardiology*, vol. 48, no. 6, pp. 933–937, 2015.

[32] M. Etemadi, O. T. Inan, R. M. Wiard, G. T. Kovacs, and L. Giovangrandi, "Non-invasive assessment of cardiac contractility on a weighing scale," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6773–6776, IEEE, 2009.

[33] M. M. H. Shandhi, B. Semiz, S. Hersek, N. Goller, F. Ayazi, and O. Inan, "Performance analysis of gyroscope and accelerometer sensors for seismocardiography-based wearable pre-ejection period estimation," *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2365–2374, 2019.

[34] F. Mosteller and J. W. Tukey, "Data analysis and regression: a second course in statistics.," *Addison-Wesley Series in Behavioral Science: Quantitative Methods*, 1977.

[35] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.

[36] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[37] A. Q. Javaid, H. Ashouri, A. Dorier, M. Etemadi, J. A. Heller, S. Roy, and O. T. Inan, "Quantifying and reducing motion artifacts in wearable seismocardiogram measurements during walking to assess left ventricular health," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 6, pp. 1277–1286, 2016.

[38] R. A. Robergs and A. F. Burnett, "Methods used to process data from indirect calorimetry and their application to vo 2 max.," *Journal of Exercise Physiology Online*, vol. 6, no. 2, 2003.

[39] M. Hedegaard, A. Anvari-Moghaddam, B. K. Jensen, C. B. Jensen, M. K. Pedersen, and A. Samani, "Prediction of energy expenditure during activities of daily living by a wearable set of inertial sensors," *Medical engineering & physics*, 2019.

[40] P. C. Mahalanobis, "On the generalized distance in statistics," National Institute of Science of India, 1936.

[41] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, ACM, 2016.

[42] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[43] T. G. Dietterich *et al.*, "Ensemble learning," *The handbook of brain theory and neural networks*, vol. 2, pp. 110–125, 2002.

[44] A. De Brabandere, T. O. De Beeck, K. H. Schütte, W. Meert, B. Vanwanseele, and J. Davis, "Data fusion of body-worn accelerometers and heart rate to predict vo2max during submaximal running," *Plos one*, vol. 13, no. 6, 2018.

[45] S. Bot and A. Hollander, "The relationship between heart rate and oxygen uptake during non-steady state exercise," *Ergonomics*, vol. 43, no. 10, pp. 1578–1592, 2000.

[46] H. Loe, Ø. Rognmo, B. Saltin, and U. Wisløff, "Aerobic capacity reference data in 3816 healthy men and women 20–90 years," *PloS one*, vol. 8, no. 5, p. e64319, 2013.

[47] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern recognition*, vol. 33, no. 1, pp. 25–41, 2000.

[48] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine learning research*, vol. 7, no. Jan, pp. 1–30, 2006.

[49] C. Vella and R. Robergs, "A review of the stroke volume response to upright exercise in healthy subjects," *British journal of sports medicine*, vol. 39, no. 4, pp. 190–195, 2005.

[50] O. T. Inan, M. Baran Pouyan, A. Q. Javaid, S. Dowling, M. Etemadi, A. Dorier, J. A. Heller, A. O. Bicen, S. Roy, T. De Marco, *et al.*, "Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients," *Circulation: Heart Failure*, vol. 11, no. 1, p. e004313, 2018.

[51] B. Londeree and S. A. Ames, "Trend analysis of the% vo2 max-hr regression.," *Medicine and science in sports*, vol. 8, no. 2, pp. 123–125, 1976.

[52] B. Londeree, T. Thomas, G. Ziogas, T. Smith, and Q. Zhang, "% vo2max versus% hrmax regressions for six modes of exercise.," *Medicine and science in sports and exercise*, vol. 27, no. 3, pp. 458–461, 1995.

[53] B. A. Franklin, J. Hodgson, and E. R. Buskirk, "Relationship between percent maximal o2 uptake and percent maximal heart rate in women," *Research quarterly for exercise and sport*, vol. 51, no. 4, pp. 616–624, 1980.

[54] H. Ashouri, S. Hersek, and O. T. Inan, "Universal pre-ejection period estimation using seismocardiography: Quantifying the effects of sensor placement and regression algorithms," *IEEE sensors journal*, vol. 18, no. 4, pp. 1665–1674, 2017.